Remarks

I. Status of the Claims

Reconsideration of this Application is respectfully requested.

Upon entry of the foregoing amendment, claims 94-95, 99-103, 105-107, and 131-146 are pending in the application, with claims 94, 107, and 139 being the independent claims. Claims 1-93, 96-98, 104, and 108-130 are canceled without prejudice to or disclaimer of the subject matter therein. Applicant reserves the right to pursue the canceled subject matter in related cases. Claims 94, 95, 100, 103, 105-107, 132-134, 136-142, and 144-146 have been amended. Amendments to claims 94, 95, 100, 132, 134, 137, 139, 140, 142, and 145 are made merely for consistency or clarification. These changes are believed to introduce no new matter, and their entry is respectfully requested.

Based on the above amendment and the following remarks, Applicant respectfully requests that the Examiner reconsider all outstanding objections and rejections and that they be withdrawn.

II. Summary of the Office Action

In the Office Action dated March 20, 2008, the Examiner has made one objection to and nine rejections of the claims. Applicant respectfully offers the following remarks concerning each of these elements of the Office Action.

III. Statement of Substance of the Interview

Applicant thanks Examiner Ford for the courtesy extended in the Telephone Interview of May 20, 2008 with Applicant's representatives, Elizabeth J. Haanes and Carla Ji Eun Kim. The Examiner's objections and rejections under 35 U.S.C. § 112, first paragraph and second paragraph were discussed. The undersigned agreed to present arguments demonstrating written description support, enablement of the claimed subject matter, and definiteness of the claims, as well as the claim amendments presented herein.

IV. Objection to Information Disclosure Statement.

The Examiner objected to the Information Disclosure Statement filed on October 31, 2007 for failing to comply with 37 C.F.R. § 1.98(a)(1). In particular, the Examiner asserted that "Applicant should submit a 1449 form along with a copy of the references to be considered." Applicant respectfully notes that PTO/SB/08A and PTO SB/08B forms have been submitted along with the Seventh Supplemental Information Disclosure Statement Pleading on October 31, 2007 as evidenced by the attached stamped postcard receipt. Courtesy copies of the PTO/SB/08A and PTO/SB/08B forms are attached. Therefore, Applicant respectfully requests that the objection be withdrawn and that the Examiner consider the cited references in the forms and execute and return the forms to Applicant.

V. Objection to Claim 94

The Examiner objected to claim 94 because it contains a period after the term "encoded." Applicant thanks the Examiner for bringing this error to Applicant's attention.

The period has been deleted, rendering the objection moot. Therefore, Applicant respectfully requests that the objection be withdrawn.

VI. Rejections Under 35 U.S.C. § 112, First Paragraph

A. Enablement of Biological Deposit

Claims 139-146 were rejected under 35 U.S.C. § 112, first paragraph, for failing to comply with the enablement requirement. In particular, the Examiner asserted that the description of the deposit of plasmid M15pReP(pQE-pmpE Ct)#37 deposited under ATCC accession no. PTA-2462 on pages 57-58 of the specification does not satisfy the requirement under 36 C.F.R. § 1.801 - § 1.809.

Applicant respectfully points to the declaration filed on December 3, 2001, attached hereto as Exhibit A. In the Declaration, Applicant stated (1) that *E.coli* containing plasmid M15 pREP (pQE-pmpE-Ct)#37 was deposited on September 12, 2000 with the American Type Tissue Culture Collection (ATCC) in compliance with the Budapest Treaty and (2) that "all restrictions on the availability to the public of a sample of the deposited microorganism will be irrevocably removed upon issuance of a United States Patent of which the microorganism are the subject." *See* Exhibit A.

Furthermore, the name of the Deposit, full street address of the depository, and the date of the Deposit were disclosed in the application as originally filed. Accordingly, Applicant respectfully notes that Applicant satisfied all requirements under 37 C.F.R. § 1.801 - § 1.809 and requests that the rejection be withdrawn.

B. Enablement for Fragments of Known Polypeptides

Claims 134 and 142 were rejected under 35 U.S.C. § 112, first paragraph for "not reasonably [providing] enablement for fragments of the *Chlamydia trachomatis* high molecular weight protein (HMW) or the *Chlamydia* major outer membrane protein (MOMP)." Office Action at page 8. Applicant respectfully traverses and disagrees with the rejection.

To be enabled, a claimed invention must be described so that any person skilled in the art can make and use the invention without undue experimentation. *In re Wands*, 858 F.2d 731, 737, 8 USPQ2d 1400, 1404 (Fed. Cir. 1988). However, the specification need not explicitly teach those in the art to make and use the invention; the requirement is satisfied if, given what they already know, the specification teaches those in the art enough that they can make and use the invention without "undue experimentation." *Amgen, Inc., v. Hoechst Marion Rousell, Inc.*, 65 U.S.P.Q.2d 1385, 1400 (Fed. Cir. 2003). "The amount of guidance or direction needed to enable the invention is inversely related to the amount of knowledge in the state of the art as well as the predictability in the art." M.P.E.P §2164.03 (Rev. 6., August 2007) (citing *In re Fisher*, 427 F.2d 833, 839, 16 USPQ 18, 24 (C.C.P.A. 1970).

Both claims 134 and 142 depend from claims 132 and 139, respectively, which are each directed to a vaccine comprising an isolated PMPE polypeptide, and further comprising one or more heterologous polypeptides. Claims 134 and 142 further require the claimed vaccine to comprise the HMW or MOMP polypeptide or fragments thereof. Due to the open-ended language "comprises," the claimed vaccine can include additional elements.

The HMW and MOMP polypeptides were well known before the filing date of this application. The HMW polypeptide and fragments thereof were disclosed in International

Publication No. WO 99/17741, filed October 01, 1998 and published April 01, 1999, prior to filing of this application. *See* page 21, lines 31-34 of the specification as originally filed. For example, the PCT publication discloses an N-terminal fragments of the HMW polypeptide such as SEQ ID NO: 17. *See* page 52, line 35 - page 53, line 2. The MOMP polypeptide and fragments thereof were also well known in the art as they were disclosed in U.S. Patent No. 5,869,608, filed March 16, 1992 and issued February 9, 1999. *See* page 22, lines 1-2 of the specification as originally filed. The '608 patent also discloses various fragments of the MOMP polypeptide including variable domains and conserved domains. An example of a highly conserved fragment of the MOMP polypeptide is a nine amino acid sequence (TTLNPTIAG) as shown at col. 8, lines 24-29. Therefore, because the HMW and MOMP polypeptide sequences as well as functional fragments were well known, their inclusion in the claimed vaccine composition would not require undue experimentation.

In view of the amendments and arguments above, Applicant respectfully requests that the Examiner reconsider and withdraw the rejection.

C. Enablement for Vaccine Compositions

Claims 94, 95, 99-103, 106-107, and 131-146 were rejected under 35 U.S.C. § 112, first paragraph, for failing to comply with the enablement requirement. While the Examiner acknowledged that the specification "[is enabled] for immunogenic compositions that produce an immune response in a subject," the Examiner asserted that the specification "does not provide enablement for [] vaccine compositions." Office Action at page 12 (emphasis in original).

The Examiner acknowledged that "[t]he instant specification has shown that there are cellular and humoral immune responses elicited when animals are administered the polypeptides of the invention." Office Action at page 15. The Examiner also noted that "[t]he specification at section 6.9 discloses in an *in vitro* neutralization model and a mouse model of salpingitis and fertility." *Id.* at page 14. However, the Examiner alleged that "[t]he specification does not provide substantive evidence that the claimed vaccines are capable of inducing protective immunity." *Id.* Applicant respectfully disagrees with these assertions.

As discussed previously, the specification need not explicitly teach those in the art to make and use the invention where the knowledge is in the art. *Amgen, Inc., v. Hoechst Marion Rousell, Inc.*, 65 U.S.P.Q.2d 1385, 1400 (Fed. Cir. 2003). The requirement is met if the specification teaches those in the art enough that they can make and use the invention without "undue experimentation. *Id.* "The key word is 'undue,' not 'experimentation.'" *Id.* at 737 (quoting *In re Angstadt*, 537 F.2d at 504).

> The test [for undue experimentation] is not merely quantitative, since a considerable amount of experimentation is permissible, if it is merely routine, or if the specification in question provides a reasonable amount of guidance with respect to the direction in which the experimentation should proceed to enable the determination of how to practice a desired embodiment of the claimed invention.

Johns Hopkins University v. Cellpro, Inc., 152 F.3d 1342 (Fed. Cir. 1998) (quoting *PPG Indus., Inc. v. Guardian Indus. Corp.*, 75 F.3d 1558, 1564 (Fed. Cir. 1996) (quotation and citation omitted) (bracketed text in original)).

In support of the claims, Applicant submitted a Supplemental Declaration under 37 C.F.R. § 1.132 on November 16, 2007. The Supplemental Declaration provided that "post filing data [] demonstrated that immunization with a vaccine comprising the serovar L2

PMPE polypeptide purified as described in Example 6.13 and 6.15 of the present application [] reduced infertility induced by *C. trachomatis* serovar F, in a standard vaginal infectivity and fertility animal models of *C. trachomatis* disease." Paragraph 18 of the Supplemental Declaration (citing U.S. Application No. 10/398,248, which claims priority to International Application No. PCT/US01/30345, which is a continuation-in-part application of the present application). Furthermore, Applicant submitted post-filing data showing the result of Example 6.9.2 as exhibit B on December 3, 2001, along with the Amendment and Reply under 37 C.F.R. § 1.111. The post-filing data, attached hereto as Exhibit B, is a table that is identical to Table 4 disclosed in Application No. 10/398,248 (Att. Dkt. No. 2479.0050001). According to the Amendment and Reply under 37 C.F.R. § 1.111 submitted on December 3, 2001, Applicant had stated that:

> [the post filing data] presents results obtained using the teaching of the specification for use of PMPE to protect against *Chlamydia* using an *in vivo* animal model. Results demonstrating the ability of PMPE to protect C3HeJOUj mice using the procedure disclosed in the specification are shown in Exhibit B. Groups of mice were immunized intranasally (i.n.) with PMPE (with or without AB5 as an adjuvant) prior to challenge with live C. trachomatis. Negative control animals were "immunized" with adjuvant alone (AB5) intranasally prior to administration of live *C. trachomatis*. Positive control animals were "immunized with adjuvant along intranasally but were not administered live *C. trachomatis*. The fertility rate for mice vaccinated with PMPE or PMPE and adjuvant (AB5) was 50% and 46% respectively. The fertility rate of mice immunized with adjuvant alone (AB5) was 9% and the fertility of mice not infected with *C. trachomatis* but administered adjuvant (AB5) was 95%. These results demonstrate that PMPE is an effective vaccine for ameliorating infertility induced by infection with *C. trachomatis*.

Pages 14-15 of the Amendment and Reply submitted on December 3, 2001. Therefore, Applicant has already submitted sufficient post-filing data demonstrating the protective immunity of the PMPE vaccine composition.

Therefore, Applicant respectfully argues that the enablement requirement under 35 U.S.C. § 112, first paragraph, has been satisfied and request that the rejection be reconsidered and withdrawn.

VII. Rejection Under 35 U.S.C. § 112, Second Paragraph

A. The term "High Molecular Weight (HMW) Protein"

The Examiner rejected claims 106, 134, and 142 under 35 U.S.C. 112, second paragraph because the Examiner alleged that "[t]he instant specification does not define 'HMW' or high molecular weight polypeptides." Office Action at page 17. Applicant respectfully disagrees with the assertion.

The specification describes the High Molecular Weight (HMW) protein by reference to a previously-published application. *See* page 21, lines 31-34 at the specification as originally filed. As indicated, the HMW protein is defined as disclosed in U.S. Patent Application Serial No. 08/942,596, filed October 2, 1997 ("the '596 application"). This information was published on April 15, 1999 in PCT Publication No. WO 99/17744, which claims priority to the '596 application. The PCT application defines the HMW polypeptide that:

> ... the HMW protein has an apparent molecular weight of about 105-115 kDa, as determined by SDS-PAGE, or a fragment or analogue thereof. Preferably the HMW protein has substantially the amino acid sequence of any of SEQ ID Nos.: 2, 15 and 16. ... It is intended that all species of *Chlamydia* are included in this invention, however preferred species include *Chlamydia trachomatis*, *Chlamydia psittaci*, *Chlamydia percorum* and *Chlamydia pneumoniae*.

Page 4, line 34 - page 5, line 16. Therefore, in view of the disclosure in the specification and knowledge in the art, Applicant respectfully requests that the rejection be reconsidered and withdrawn.

B. The term "Mature Putative Membrane Protein E"

The Examiner rejected claims 107 and 131-133 because the Examiner alleged that the structure of the mature putative membrane protein, *e.g.*, amino acids sequence, is not clear. *See* Office Action at page 17. Applicant respectfully traverses the rejection.

In reviewing a claim for compliance with 35 U.S.C. § 112, second paragraph, the Examiner must consider the claim as a whole to determine whether the claim apprises one of ordinary skill in the art of its scope. *See Solomon v. Kimberly Clark Corp.*, 216 F.3d 1372, 1379, 55 USPQ2d 1279, 1283 (Fed. Cir. 2000). The principle inquiry under 35 USC § 112, second paragraph, is whether a person of ordinary skill in the art would be apprised of the scope of the claim. *See Solomon v. Kimberly-Clark Corp.*, 216 F.3d 1372, 1379, 55 USPQ2d 1279, 1283 (Fed. Cir. 2000).

As pointed out during the interview held on May 20, 2008, Applicant has already pointed out in the Amendment and Reply under 37 C.F.R. § 1.114, filed on October 31, 2007, that the scope of "the mature putative membrane protein E" would have been easily predictable by a skilled artisan. A broad base of scientific and technical knowledge relating to secreted proteins has been in place for at least 25 years. *See e.g.*, Lewin, B., *Genes*, John Wiley and Sons (1983), at pp. 159-160 (attached hereto as Exhibit C). Even in 1983 it was well understood that "the N-terminus of secreted proteins consists of a cleavable leader of from 16-29 amino acids, which starts with two or three polar residues, but continues with a

high content of hydrophobic amino acids . . . ." *Id.* at 159. As is well known in the art, once a signal peptide is cleaved, what is left is the "mature" polypeptide. The knowledge since 1983 has become increasingly more sophisticated such that signal peptides can be easily predicted based on the amino acid structure. Indeed, an internet-based algorithm for predicting signal peptides of secreted proteins was available in 1997. *See, e.g.*, Nielsen *et al. Protein Engineering 10*:1-6 (1997) at page 5 (attached hereto as Exhibit D). The most recent version of this program may be found at www.cbs.dtu.dk/services/SignalP/ (last visited June 20, 2008). Therefore, in view of the knowledge available in the art, Applicant respectfully argues that the term "mature putative membrane protein E" is clear and definite and request that the rejection be reconsidered and withdrawn.

C. The terms "Pre or Pro-Sequence" and "Immunogenic Sequence"

The Examiner rejected claims 107, 131- 133, and 141-144 under 35 U.S.C. § 112, second paragraph because the Examiner alleged that "the structure [or amino acids] of the claimed pre-sequence, pro-sequence or immunogenic sequence" are required. *Id.* at page 18. Applicant respectfully traverses and disagrees with the rejection.

Applicant respectfully notes that only claims 133 and 141, but not claims 107, 131, 132, and 142-144, recite the terms "pre or pro sequence" and "immunogenic sequence." Furthermore, one skilled in the art could easily ascertain the scope of the claims. First, the terms "pre or pro sequence" and "immunogenic sequence" are explained in the specification at page 21, lines 24-29 that:

> [u]seful heterologous polypeptides to be included within such a chimeric polypeptide include, but are not limited to, a) pre- and/or pro-sequences that facilitate the transport, translocation and/or processing of the PMP-derived polypeptide in a host cell, b) affinity purification sequences, and c) any useful

immunogenic sequences (e.g., sequences encoding one or more epitopes of a surface-exposed protein of a microbial pathogen).

As shown in the specification, pre- and/or pro-sequences that facilitate the transport, translocation and/or processing of the PMP-derived polypeptide in a host cell are well known in the art. Furthermore, the specification explains the term "immunogenic sequence" and provides examples of the immunogenic sequences at page 21, line 24 to page 22, line 2. Applicant argues that the requirement under 35 U.S.C. § 112, second paragraph is to define the boundaries of the subject matter for which protection is sought, but not to disclose every possible embodiment of the claims. In light of the disclosure in the specification and knowledge available in the art, a person of ordinary skill in the art would readily ascertain the scope of the claims. Therefore, Applicant respectfully requests that the rejection be reconsidered and withdrawn.

D. Trademark "Ribi DETOX"

Claims 136 and 144 were rejected under 35 U.S.C. § 112, second paragraph for being indefinite. Applicant has deleted the trademark in the claims, rendering the rejection moot. Applicant respectfully requests that the rejection be reconsidered and withdrawn. In addition, claims 136 and 144 have been amended to add additional adjuvants. Support for these amendments may be found at page 50, line 29 to page 51, line 1.

E. "PmpE encoded by SEQ ID NO: 2"

The Examiner rejected claims 107 and 131-138 were rejected under 35 U.S.C. § 112, second paragraph," alleging that "SEQ ID NO: 2 is an amino acid sequence and cannot

encode a polypeptide." *Id.* Applicant has amended the claim by replacing "encoded by" to "contained in." Support for the claim amendment can be found at page 10, lines 5-9 at the specification as originally filed. Therefore, the rejection is rendered moot, and Applicant respectfully requests that the rejection be withdrawn.

F. The term "Bacterial Toxin or Fragment thereof"

The Examiner rejected claim 138 was rejected under 35 U.S.C. § 112, alleging that "it is unclear as to what Applicant intends by a fragment of a bacterial toxin." Applicant respectfully traverses the rejection. Nonetheless, in order to advance prosecution of this application, but not to acquiesce the Examiner's rejection, Applicant has deleted the "fragment" language from claim 138, rendering the rejection moot. Applicant respectfully requests that the rejection be reconsidered and withdrawn.

Conclusion

All of the stated grounds of objection and rejections have been properly traversed, accommodated, or rendered moot. Applicant therefore respectfully requests that the Examiner reconsider all presently outstanding objections and rejections and that they be withdrawn. Applicant believes that a full and complete reply has been made to the outstanding Office Action and, as such, the present application is in condition for allowance. If the Examiner believes, for any reason, that personal communication will expedite prosecution of this application, the Examiner is invited to telephone the undersigned at the number provided.

Prompt and favorable consideration of this Amendment and Reply is respectfully requested.

Respectfully submitted,

STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

Reg # 36,688

Elizabeth J. Haanes, Ph.D.
Attorney for Applicant
Registration No. 42,613

Date: 6/20/01

1100 New York Avenue, N.W.
Washington, D.C. 20005-3934
(202) 371-2600

811382v1

EXHIBIT A

EXHIBIT A

Express Mail No.: EL 477 032 898 US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of: W. James Jackson

Serial No.: 09/677,752 Group Art Unit: 1645

Filed: October 3, 2000 Examiner: V. Ford

For: CHLAMYDIA PROTEIN, GENE SEQUENCE AND USES THEREOF Attorney Docket No.: 7969-087-999

STATEMENT REGARDING PERMANENCE AND AVAILABILITY OF DEPOSITED MICROORGANISMS

Assistant Commissioner for Patents
U.S. Patent and Trademark Office
P.O. Box 2327
Arlington, VA 22202

Sir:

I, W. James Jackson, declare and state:

1. That I am an authorized Officer of Antex Biologics Inc., the Assignee of the above-identified application.

2. That on September 12, 2000, *E. coli* containing plasmid M15 pREP (pQE-pmpE-Ct) #37 was deposited with the AMERICAN TYPE TISSUE CULTURE COLLECTION (ATCC), at 10801 UNIVERSITY BLVD., MANASSAS, VIRGINIA 20110-2209, USA, International Depository Authority, in compliance with the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure. The deposit was viable at the time of deposit and has been assigned accession number ATCC No. PTA-2462.

3. That I hereby assure the United States Patent and Trademark Office and the public that (a) all restrictions on the availability to the public of a sample of the

deposited microorganism will be irrevocably removed upon issuance of a United States patent of which the microorganism are the subject; (b) the above-mentioned microorganism will be maintained for a period of at least five years after the most recent request for the furnishing of a sample of the deposited microorganism were received by the ATCC and, in any case for a period of at least 30 years after the date of deposit; (c) should the deposited microorganism become non-viable it will be replaced by the Assignee; and (d) access to the deposited microorganism will be available to the Commissioner during the pendency of the patent application or to one determined by the Commissioner to be entitled to such cell line under 37 C.F.R. § 1.14 and 35 U.S.C. § 122.

I hereby declare further that all statements made herein of my own knowledge are true and all statements made on information and belief are believed to be true and further I make these statements with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under §1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Respectfully submitted,

Date: 11/29/01

By: _____

W. James Jackson, Vice President
Antex Biologics Inc.

EXHIBIT B

Fertility Assessment for pmpE (FL / GP)

Group	Vaccine & Route	Fertile Females per Total	% Fertility	Number Litters per Total
I	PmpE + AB5 / i.n.	4 / 8	50%	5 / 8
II	PmpE / i.n.	7 / 15	46%	8 / 15
III	AB5 / i.n. (Neg. Control)	2 / 22	9%	3 / 22
IV	AB5 / i.n. (Pos. Control)	19 / 20	95%	41 / 20

EXHIBIT C

EXHIBIT C

GENES

BENJAMIN LEWIN
Editor, *Cell*

John Wiley & Sons
New York Chichester Brisbane
Toronto Singapore

Production supervised by Linda Indig
Manuscript editor was Mondy Dana,
supervised by Vivian Kahane Cooper
Cover and text designed by Laura Ierardi
Illustrations by John Balbalis with
the assistance of the Wiley
illustration department.

Library of Congress Cataloging in Publication Data:

Lewin, Benjamin
Genes.

Includes bibliographies and indexes.
1. Genetics. I. Title.

QH430.L487 1983 575.1 82-20022
ISBN 0-471-09316-5

Printed in the United States of America

10 9 8 7 6 5 4 3 2

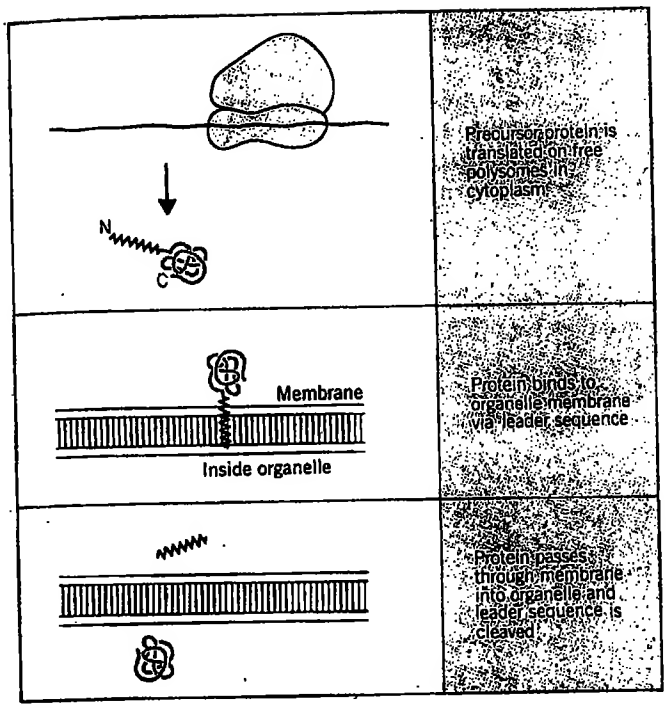

Figure 9.12
Leader sequences are used for proteins to recognize mitochondrial or chloroplast surfaces.

For many proteins that must be inserted in membranes, the sequence of the mature polypeptide is not itself sufficient to direct membrane insertion. Additional information is needed; and this most often takes the form of a **leader sequence** at the N-terminal end of the protein. The protein carrying this leader is called a **preprotein.** It is a transient precursor to the mature protein, since the leader is cleaved as part of the process of membrane insertion.

The **pre** sequence is distinct from the **pro** sequence that describes the additional regions present on proteins that exist as *stable* precursors. Some proteins may have both. For example, insulin is initially synthesized as **preproinsulin;** the **pre** sequence is cleaved during secretion, generating **proinsulin,** which is the substrate for processing to mature insulin.

The leader sequence plays different roles in different circumstances. For certain proteins synthesized within the cytoplasm, but destined to reside within the chloroplast or mitochondrion, the product of cytoplasmic protein synthesis is a precursor some 5000 daltons (roughly 45 amino acids) larger than the mature protein. This precursor is released from polysomes. If it is added to intact organelles *in vitro*, it can be incorporated into the compartment. As illustrated in **Figure 9.12,** this involves passage through the organelle membrane, during which the leader sequence is cleaved, probably by a protease located on the outside of the envelope. The leader sequence serves to provide information recognized by the organelle membrane and used to sequester the protein in a **post-translational** process. Note that a cleavable leader is not the only acceptable form of such information; some mitochondrial proteins are recognized as such in their mature form, and may have an internal sequence that is able to ensure membrane passage without cleavage.

For proteins that are secreted through, or inserted into, other cellular membranes, the process of association most often starts during translation. The polyribosomes synthesizing these proteins are associated with the membrane of the endoplasmic reticulum. The preproteins are not released into the cytoplasm to form a precursor pool, but instead pass directly from the ribosome to the membrane. From the membrane, the proteins enter the Golgi apparatus, and then are directed to their ultimate destination, such as the lysosome or the plasma membrane.

A model for the mechanism of membrane insertion has been based on work with eucaryotic microsomal systems (containing ribosomes and endoplasmic reticulum). These systems are able to package *nascent* proteins into membranes; but they do not work with the addition of isolated preproteins. The **signal hypothesis** proposes that the leader characteristic of almost all secreted proteins constitutes a **signal sequence** whose presence distinguishes them from other proteins. With only rare exceptions, the N-terminus of secreted proteins consists of a cleavable leader of from 16 to 29 amino acids, which starts with two or three polar residues, but continues with a high content of hydrophobic amino acids; otherwise there is no noticeable conservation of sequence.

The signal sequence provides the means for ribosomes translating the mRNA to attach to the membrane. Some membrane receptor recognizes the signal sequence, perhaps by virtue of its hydrophobicity,

and inserts the precursor protein directly into the membrane, probably as soon as the signal sequence and a few additional amino acids have been synthesized. A route to characterizing the protein receptors is provided by the discovery that salt-washed membranes cannot sponsor ribosomal attachment; but this ability can be recovered by adding the salt wash. The active component has been purified in the form of a complex of six proteins.

Figure 9.13 shows that as synthesis of the nascent polypeptide chain continues, there comes a point at which the protein is well inserted into the membrane, and the signal sequence can be cleaved. Then when the ribosomes complete translation, the protein is al-

Figure 9.13
The signal hypothesis proposes that ribosomes synthesizing secretory proteins are attached to the membrane via the leader sequence on the nascent polypeptide.

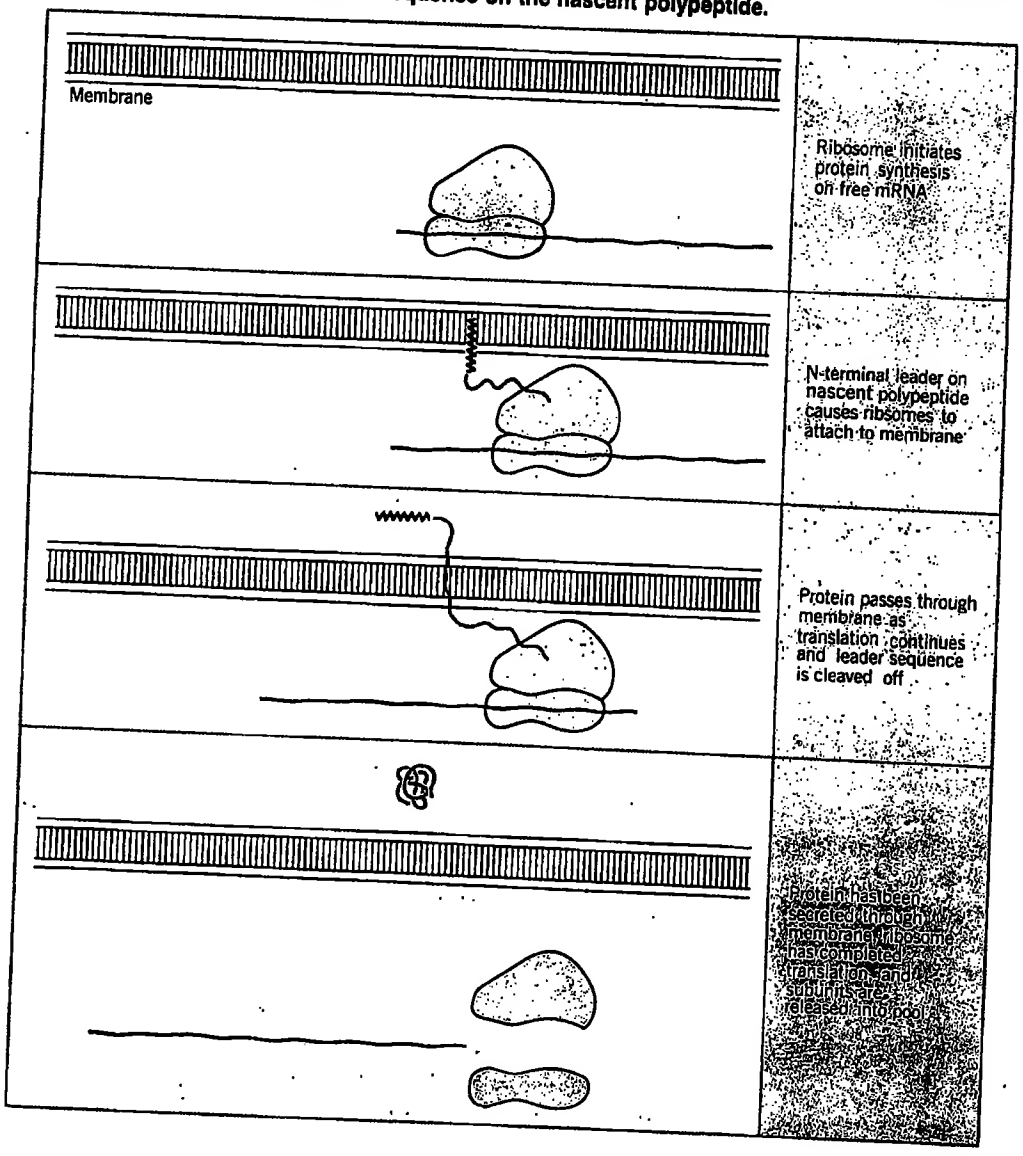

EXHIBIT D

Protein Engineering vol.10 no.1 pp.1–6, 1997

SHORT COMMUNICATION

Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites

Henrik Nielsen, Jacob Engelbrecht[1], Søren Brunak and Gunnar von Heijne[2]

Center for Biological Sequence Analysis, Department of Chemistry, The Technical University of Denmark, DK-2800 Lyngby, Denmark and [2]Department of Biochemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden

[1]Present address: Novo Nordisk A/S, Scientific Computing, Building 9M1, Novo Alle, DK-2880 Bagsværd, Denmark

We have developed a new method for the identification of signal peptides and their cleavage sites based on neural networks trained on separate sets of prokaryotic and eukaryotic sequence. The method performs significantly better than previous prediction schemes and can easily be applied on genome-wide data sets. Discrimination between cleaved signal peptides and uncleaved N-terminal signal-anchor sequences is also possible, though with lower precision. Predictions can be made on a publicly available WWW server.

Keywords: cleavage sites/protein sorting/secretion/signal peptide

Introduction

Signal peptides control the entry of virtually all proteins to the secretory pathway, both in eukaryotes and prokaryotes (Gierasch, 1989; von Heijne, 1990; Rapoport, 1992). They comprise the N-terminal part of the amino acid chain and are cleaved off while the protein is translocated through the membrane. The common structure of signal peptides from various proteins is commonly described as a positively charged n-region, followed by a hydrophobic h-region and a neutral but polar c-region. The (–3,–1) rule states that the residues at positions −3 and −1 (relative to the cleavage site) must be small and neutral for cleavage to occur correctly (von Heijne, 1983, 1985).

A strong interest in the automated identification of signal peptides and the prediction of their cleavage sites has been evoked not only by the huge amount of unprocessed data available, but also by the industrial need to find more effective vehicles for the production of proteins in recombinant systems. The most widely used method for predicting the location of the cleavage site is a weight matrix which was published in 1986 (von Heijne, 1986). This method is also useful for discriminating between signal peptides and non-signal peptides by using the maximum cleavage site score. The original matrices are commonly used today, even though the amount of signal peptide data available has increased since 1986 by a factor of 5–10.

Here, we present a combined neural network approach to the recognition of signal peptides and their cleavage sites, using one network to recognize the cleavage site and another network to distinguish between signal peptides and non-signal peptides. A similar combination of two pairs of networks has been used with success to predict the intron splice sites in pre-mRNA from humans and the dicotelydoneous plant *Arabidopsis thaliana* (Brunak *et al.*, 1991; S.Hebsgaard, P.Korning, J.Engelbrecht, P.Rouźe and S.Brunak, submitted). Artificial neural networks have been used for many biological sequence analysis problems (Hirst and Sternberg, 1992; Presnell and Cohen, 1993). They have also been applied to the twin problems of predicting signal peptides and their cleavage sites, but until now without leading to practically applicable prediction methods with significant improvements in performance compared with the weight matrix method (Arrigo *et al.*, 1991; Ladunga *et al.*, 1991; Schneider and Wrede, 1993).

Materials and methods

The data were taken from SWISS-PROT version 29 (Bairoch and Boeckmann, 1994). The data sets were divided into prokaryotic and eukaryotic entries and the prokaryotic data sets were further divided into Gram-positive eubacteria (*Firmicutes*) and Gram-negative eubacteria (*Gracilicutes*), excluding *Mycoplasma* and *Archaebacteria*. Viral, phage and organellar proteins were not included. In addition, two single-species data sets were selected, a human subset of the eukaryotic data and an *Escherichia coli* subset of the Gram-negative data.

The sequence of the signal peptide and the first 30 amino acids of the mature protein from the secretory protein were included in the data set. The first 70 amino acids of each sequence were used from the cytoplasmic and (for the eukaryotes) nuclear proteins. In addition, a set of eukaryotic signal anchor sequences, i.e. N-terminal parts of type II membrane proteins (von Heijne, 1988), were extracted (see Figure 1).

As an example of a large-scale application of the finished method, we used the *Haemophilus influenzae* Rd genome—the first genome of a free-living organism to be completed (Fleischmann *et al.*, 1995). We have downloaded the sequences of all the predicted coding regions in the *H.influenzae* genome from the World Wide Web (WWW) server of the Institute for Genomic Research at http://www.tigr.org/. Only the first 60 positions of each sequence were analysed.

We have attempted to avoid signal peptides where the cleavage sites are not experimentally determined, but we are not able to eliminate them completely, since many database entries simply lack information about the quality of the evidence. The details of the data selection are described in the WWW server and in an earlier paper (Nielsen *et al.*, 1996a).

Redundancy in the data sets was avoided by excluding pairs of sequences which were functionally homologous, i.e. those that had more than 17 (eukaryotes) or 21 (prokaryotes) exact matches in a local alignment (Nielsen *et al.*, 1996a). Redundant sequences were removed using an algorithm which guarantees that no pairs of homologous sequences remain in the data set (Hobohm *et al.*, 1992). This procedure removed 13–56% of the sequences. The numbers of non-homologous sequences remaining in the data sets are shown in Table I. Redundancy

Table I. Data and performance values

Source	Data (Number of sequences)		Network architecture (window/hidden units)		Performance	
	Signal peptides	Non-secretory proteins	*C*-score	*S*-score	Cleavage site location (% correct)	Signal peptide discrimination (correlation)
Human	416	251	15+4/2	27 / 4	68.0 (67.9)	0.96 (0.97)
Eukaryote	1011	820	17+2/2	27 / 4	70.2	0.97
E.coli	105	119	15+2/2	39 / 0	83.7 (85.7)	0.89 (0.92)
Gram–	266	186	11+2/2	19 / 3	79.3	0.88
Gram+	141	64	21+2/0	19 / 3	67.9	0.96

Data: the number of sequences of signal peptides and non-secretory (i.e. cytoplasmic or nuclear) proteins in the data sets after redundancy reduction. The organism groups are eukaryotes, human, Gram-negative bacteria ('Gram–'), *E.coli* and Gram-positive bacteria ('Gram+'). The human data are subsets of the eukaryotic data and the *E.coli* data are subsets of the Gram-negative data. The signal anchor and *H.influenzae* data are not shown in the table. *Network architecture*: the size of the input window and the number of hidden computational units ('neurons') in the optimal neural networks chosen for each data set. *C*-score networks have asymmetrical input windows. *Performance*: the percentage of signal peptide sequences where the cleavage site was predicted to be at the correct location according to the maximal value of the Y-score (see Figure 2). The ability of the method to distinguish between the signal peptides and the N-terminals of non-secretory proteins (based on the mean value of the *S*-score in the region between position 1 and the predicted cleavage site position) is measured by the correlation coefficients (Mathews, 1975). Both performance values are measured on the test sets (the average of five cross-validation tests). The values given in parentheses indicate the performance for the human sequences when using networks trained on all eukaryotic data and for the *E.coli* sequences when using Gram-negative networks respectively.

reduction was not applied to the signal anchor data or the *H.influenzae* data, since these were not used as training data.

Neural network algorithms

The signal peptide problem was posed to the neural networks in two ways: (i) recognition of the cleavage sites against the background of all other sequence positions and (ii) classification of amino acids as belonging to the signal peptide or not. In the latter case, negative examples included both the first 70 positions of non-secretory proteins and the first 30 positions of the mature part of secretory proteins.

The neural networks were feed-forward networks with zero or one layer of two to 10 hidden units, trained using back-propagation (Rumelhart *et al.*, 1986) with a slightly modified error function. The sequence data were presented to the network using sparsely encoded moving windows (Qian and Sejnowski, 1988; Brunak *et al.*, 1991). Symmetric and asymmetric windows of a size varying from five to 39 positions were tested.

Based on the numbers of correctly and incorrectly predicted positive and negative examples, we calculated the correlation coefficient (Mathews, 1975). The correlation coefficients of both the training and test sets were monitored during training and the performance of the training cycle with the maximal test set correlation was recorded for each training run. The networks chosen for inclusion in the WWW server have been trained until this cycle only.

The test performances have been calculated by cross-validation: each data set was divided into five approximately equal-sized parts and then every network run was carried out with one part as test data and the other four parts as training data. The performance measures were then calculated as an average over the five different data set divisions.

For each of the five data sets, one signal peptide/non-signal peptide network architecture and one cleavage site/non-cleavage site network architecture was chosen on the basis of the test set correlation coefficients. We did not pick the architecture with absolutely the best performance, but instead the smallest network that could not be significantly improved by enlarging the input window or adding more hidden units.

The trained networks provide two different scores between zero and one for each position in an amino acid sequence. The output from the signal peptide/non-signal peptide networks, the *S*-score, can be interpreted as an estimate of the probability of the position belonging to the signal peptide, while the output from the cleavage site/non-cleavage site networks, the *C*-score, can be interpreted as an estimate of the probability of the position being the first in the mature protein (position +1 relative to the cleavage site).

If there are several *C*-score peaks of comparable strength, the true cleavage site may often be found by inspecting the *S*-score curve in order to see which of the *C*-score peaks coincides best with the transition from the signal peptide to the non-signal peptide region. In order to formalize this and improve the prediction, we have tried a number of linear and non-linear combinations of the raw network scores and evaluated the percentage of sequences with correctly placed cleavage sites in the five test sets. The best measure was the geometric average of the C-score and a smoothed derivative of the S-score, termed the Y-score:

$$Y_i = \sqrt{C_i \Delta_d S_i}, \tag{1}$$

where $\Delta_d S_i$ is the difference between the average *S*-score of d positions before and d position after position i:

$$\Delta_d S_i = \frac{1}{d}\left(\sum_{j=1}^{d} S_{i-j} - \sum_{j=0}^{d-1} S_{i+j} \right) \tag{2}$$

In Figure 2(A), examples of the values of the C-, S- and Y-scores are shown for a typical signal peptide with a typical cleavage site. The C-score has one sharp peak that corresponds to an abrupt change in the S-score from a high to low value. Among the real examples, the C-score may exhibit several peaks and the S-score may fluctuate. We define a cleavage site as being correctly located if the true cleavage site position corresponds to the maximal Y-score (combined score).

For a typical non-secretory position, the values of the C-, S- and Y-scores are lower, as shown in Figure 2(B). We found the best discriminator between signal peptides and non-secretory

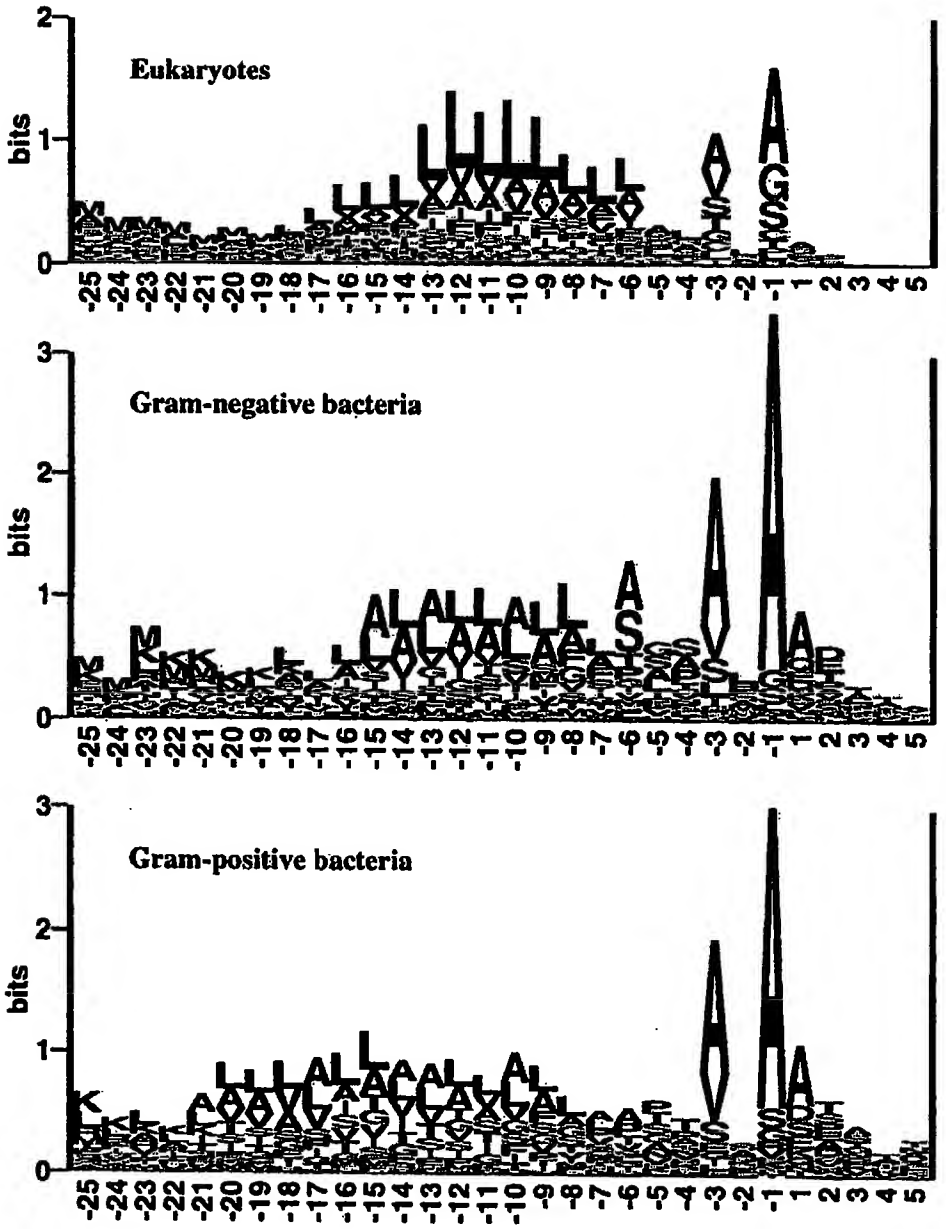

Fig. 1. Sequence logos (Schneider and Stephens, 1990) of signal peptides, aligned by their cleavage sites. The total height of the stack of letters at each position shows the amount of information, while the relative height of each letter shows the relative abundance of the corresponding amino acid. The information is defined as the difference between the maximal and actual entropy (Shannon, 1948): $I_j = H_{max} - H_j = \log_2 20 + \Sigma_\alpha\, n_j(\alpha)/N_j \log_2 n_j(\alpha)/N_j$, where $n_j(\alpha)$ is the number of occurrences of the amino acid α and N_j is the total number of letters (occupied positions) at position j. Positively and negatively charged residues are shown in blue and red respectively, while uncharged polar residues are green and hydrophobic residues are black.

proteins to be the average of the S-score in the predicted signal peptide region, i.e. from position 1 to the position immediately before the position where the Y-score has a maximal value. If this value—the mean S-score—is greater than 0.5, we predict the sequence in question to be a signal peptide (cf. Figure 3).

The relationship between the various performance measures and their development during the training process is described in detail elsewhere (Nielsen *et al.*, 1997).

Results and discussion

The optimal network architecture and corresponding predictive performance for all the data sets are shown in Table I. The C-score problem is best solved by networks with asymmetric windows, i.e. windows including more positions upstream than downstream of the cleavage site. This corresponds well with the location of the cleavage site pattern information which is shown as sequence logos (Schneider and Stephens, 1990) in Figure 1. The S-score problem, on the other hand, is best solved by symmetric or approximately symmetric windows.

Although our method is able to locate cleavage sites and discriminate signal peptides from non-secretory proteins with a reasonably high reliability, the accuracy of the cleavage site location is lower than that reported for the original weight matrix method (von Heijne, 1986): 78% for eukaryotes and

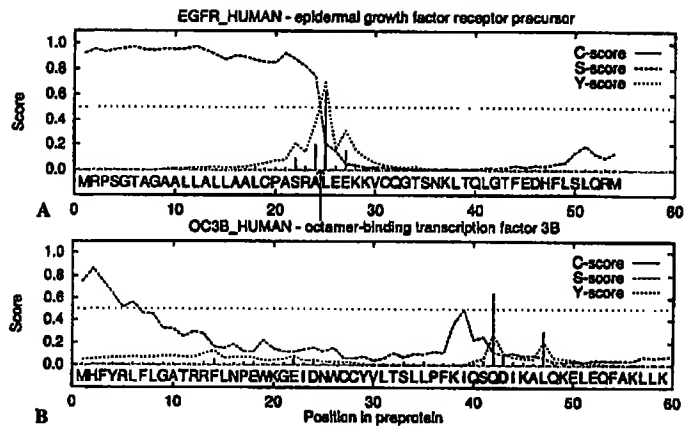

Fig. 2. Examples of network output. The values of the C- (output from cleavage site networks), S- (output from signal peptide networks) and Y-scores (combined cleavage site score, $Y_i = \sqrt{C_i \Delta_d S_i}$) are shown for each position in the sequence. The C- and S-scores are averages over five networks trained on different parts of the data. Note: the C- and Y-scores are high for the position immediately after the cleavage site, i.e. the first position in the mature protein. (A) A successfully predicted signal peptide. The true cleavage site is marked wih an arrow. (B) A non-secretory protein. For many non-secretory proteins, all three scores are very low throughout the sequence. In this example, there are peaks of the C- and S-scores, but the sequence is still easily classified as non-secretory, since the C-score peak occurs far away from the S-score decline and the region of the high S-score is far too short.

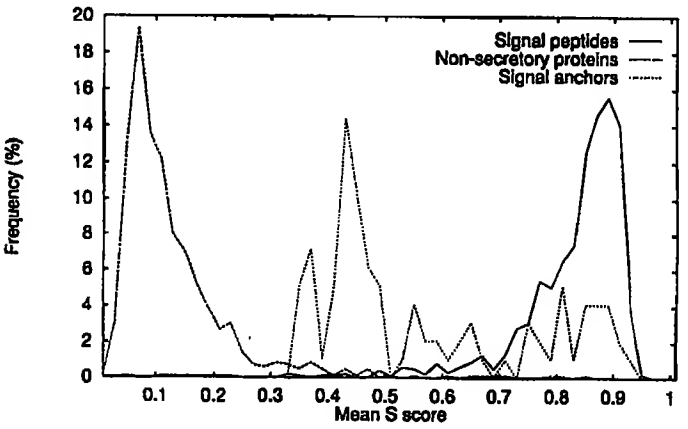

Fig. 3. Distribution of the mean signal peptide score (S-score) for signal peptides and non-signal peptides (eukaryotic data only). 'Non-secretory proteins' refer to the N-terminal parts of cytoplasmic or nuclear proteins, while 'signal anchors' are the N-terminal parts of type II membrane proteins. The mean S-score of a sequence is the average of the S-score over all positions in the predicted signal peptide region (i.e. from the N-terminal to the position immediately before the maximum of the Y-score). The bin size of the distribution is 0.02.

89% for prokaryotes (not divided into Gram-positive and -negative). When the original weight matrix is applied to our recent data set, however, the performance is much lower. This suggests a larger variation in the examples of the signal peptides found since then. It may, of course, also reflect a higher occurrence of errors in our automatically selected data than in the manually selected 1986 set.

In order to compare the strength of the neural network approach to the weight matrix method, we recalculated new weight matrices from our new data and tested the performances of these (results not shown). The weight matrix method was comparable to the neural networks when calculating the C-score, but was practically unable to solve the S-score problem

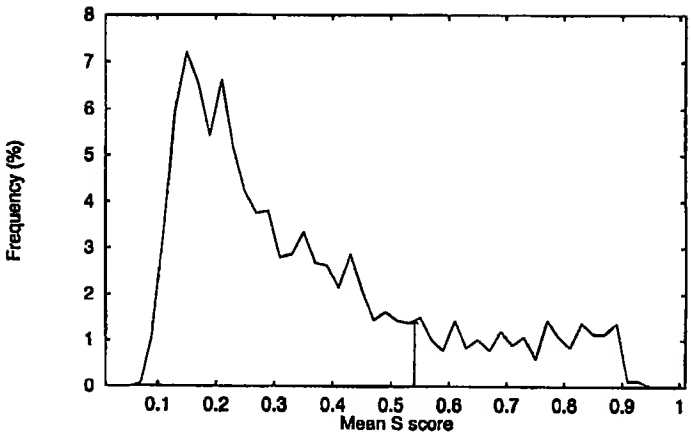

Fig. 4. Distribution of the mean signal peptide score (S-score) for all the predicted *H.influenzae* coding sequences. The mean S-score is calculated using networks trained on the Gram-negative data set. The bin size of the distribution is 0.02. The arrow shows the optimal cut-off for predicting a cleavable signal peptide. The predicted number of secretory proteins in *H.influenzae* (corresponding to the area under the curve to the right of the arrow) is 330 out of 1680 (20%).

and therefore did not provide the possibility of calculating the combined Y-score.

Note that the prediction performances reported here correspond to minimal values. The test sets in the cross-validation have a very low sequence similarity; in fact, the sequence similarity is so low that the correct cleavage sites cannot be found by alignment (Nielsen *et al.*, 1996a). This means that the prediction accuracy on sequences with some similarity to the sequences in the data sets will in general be higher.

The differences between the signal peptides from different organisms are apparent from Figure 1. The signal peptides from Gram-positive bacteria are considerably longer than those of other organisms, with much more extended h-regions, as observed previously (von Heijne and Abrahmsén, 1989). The prokaryotic h-regions are dominated by Leu (L) and Ala (A) in approximately equal proportions and in the eukaryotes they are dominated by Leu with some occurrence of Val (V), Ala, Phe (F) and Ile (I). Close to the cleavage site, the (–3,–1) rule is clearly visible for all three data sets, but while a number of different amino acids are accepted in the eukaryotes, the prokaryotes accept alanine almost exclusively in these two positions. In the first few positions of the mature protein (downstream of the cleavage site) the prokaryotes show certain preferences for Ala, negatively charged (D or E) amino acids, and hydroxy amino acids (S or T), while no pattern can be seen for the eukaryotes. In the leftmost part of the alignment, the positively charged residue Lys (K) [and to a smaller extent Arg (R)] is seen in the prokaryotes, while the eukaryotes show a somewhat weaker occurrence of Arg (barely visible in the figure) and almost no Lys. This corresponds well with the hypothesis that positive residues are required in the n-region where the N-terminal Met is formulated for prokaryotes, but not necessarily for eukaryotes where the N-terminal Met in itself carries a positive charge (von Heijne, 1985).

The difference in structure is reflected in the performances of the trained neural networks (see Table I). Gram-negative cleavage sites have the strongest pattern—i.e. the highest information content—and, consequently, they are the easiest to predict, both at the single-position and at the sequence level. The eukaryotic cleavage sites are significantly more difficult

to predict. Gram-positive cleavage sites are slightly more difficult to predict than the eukaryotic ones, which would not be expected from the sequence logos (Figure 1), since they show nearly as high an information content as the Gram-negative cleavage sites, but the longer Gram-positive signal peptides means that the cleavage sites have to be located against a larger background of non-cleavage site positions. The discrimination of signal peptides versus non-secretory proteins, on the other hand, is better for the eukaryotes than for the prokaryotes. This may be due to the more characteristic leucine-rich h-regions of the eukaryotic signal peptides.

The logos for the human and *E.coli* data sets are not shown, since they show no significant differences from those of the eukaryotes or Gram-negative bacteria respectively. Accordingly, the predictive performance was not improved by training the networks on single-species data sets. On the contrary, the *E.coli* signal peptides are predicted even better by the Gram-negative networks than by the *E.coli* networks (probably due to the relatively small size of the *E.coli* data set). In other words, we have found no evidence for species-specific features of the signal peptides of humans and *E.coli*.

Signal anchors often have sites similar to signal peptide cleavage sites after their hydrophobic (transmembrane) region. Therefore, a prediction method can easily be expected to mistake signal anchors for peptides. In Figure 3, the distribution of the mean S-score for the 97 eukaryotic signal anchors is included. It shows some overlap with the signal peptide distribution. If the standard cut-off of 0.5 is applied to the signal anchor data sets, 50% of the eukaryotic signal anchor sequences are falsely predicted as signal peptides (the corresponding figure for the human signal anchors is 75% when using human networks and 68% when using eukaryotic networks). With a cut-off optimized for signal anchor versus signal peptide discrimination (0.62), we were able to lower this error rate to 45% for the eukaryotic data set. The mean S-score still gives a better separation than the maximal C- or Y-score, which indicates that the pseudo-cleavage sites are in fact rather strong.

However, the pseudo-cleavage sites often occur further from the N-terminal than genuine cleavage sites do. If we do not accept signal peptides longer than 35 residues (this will exclude only 2.2% of the eukaryotic signal peptides in our data set), the percentage of false positives among the signal anchors drops to 28% for the eukaryotic and 32% for the human signal anchors (39% when using eukaryotic networks). When taking this into account, our method does provide a reasonably good discrimination between signal peptides and signal anchors. This has not been reported by any of the earlier published methods for signal peptide recognition.

Scanning the Haemophilus influenzae genome

We have applied the prediction method with networks trained on the Gram-negative data set to all the amino acid sequences of the predicted coding regions in the *Haemophilus influenzae* genome. The distribution of the mean S-score (from position 1 to the position with a maximal Y-score) is shown in Figure 4.

When applying the optimal cut-off value found for the Gram-negative data set, we obtained a crude estimate of the number of sequences with cleavable signal peptides in *H.influenzae*: 330 out of 1680 sequences or approximately 20%. If the maximal S-score is used instead of the mean S-score, the estimate comes out as 28% and with the maximal Y-score it is 14% (distributions not shown). If all three criteria are applied together, leaving only 'typical' signal peptides, we obtain 188 sequences (11%).

Some of the sequences predicted to be signal peptides according to the S-score but not according to the Y-score may be signal anchor-like sequences of type II (single-spanning) or type IV (multispanning) membrane proteins. This hypothesis is strengthened by a hydrophobicity analysis of the ambiguous examples (results not shown). If we apply the slightly higher cut-off optimized for the discrimination of signal anchors versus signal peptides in eukaryotes (0.62) to the mean S-score, the estimate is lowered from 20 to 15%.

On the other hand, some of the sequences predicted to be signal peptides according to the maximal Y-score but not the mean S-score may be the effect of the initiation codon of the predicted coding region having been placed too far upstream. In this case, the apparent signal peptide becomes too long and the region between the false and the true initiation codon will probably not have signal peptide character, thereby bringing the mean S-score of the erroneously extended signal peptide region below the cut-off. This is strengthened by the finding that these ambiguous examples are longer than average and contain more methionines.

In conclusion, we estimate that 15–20% of the *H.influenzae* proteins are secretory. However, a whole-genome analysis like this would be more reliable if combined with other analyses, notably transmembrane segment predictions and initiation site predictions.

Method and data publicly available

The finished prediction method is available both via an e-mail server and a WWW server. Users may submit their own amino acid sequences in order to predict whether the sequence is a signal peptide and, if so, where it will be cleaved. We recommend that only the N-terminal part (say 50–70 amino acids) of the sequences is submitted, so that the interpretation of the output is not obscured by false positives further downstream in the protein.

The user is asked to choose between the network ensembles trained on data from Gram-positive, Gram-negative or eukaryotic organisms. We did not include the networks trained on the single-species data sets in the servers, since these did not improve the performance.

The values of the C-, S- and Y-scores are returned for every position in the submitted sequence. In addition, the maximal Y-score, maximal S-score and mean S-score values are given for the entire sequence and compared with the appropriate cut-offs. If the sequence is predicted to be a signal peptide, the position with the maximal Y-score is mentioned as the most likely cleavage site. A graphical plot in postscript format, similar to those in Figure 2, may be requested from the servers. We strongly recommend that a graphical plot is always used for the interpretation of the output. The plot may give hints about, for example, multiple cleavage sites or erroneously assigned initiation, which would not be found when using only the maximal or mean score values.

The address of the mail server is signalp@cbs.dtu.dk. For detailed instructions, send a mail containing the word 'help' only. The WWW server is accessible via the Center for Biological Sequence Analysis homepage at http://www.cbs.dtu.dk/.

All the data sets mentioned in Table I are available from an FTP server at ftp://virus.cbs.dtu.dk/pub/signalp. Retrieve the file README for detailed descriptions of the data and the format.

The FTP server and the mail server can both be accessed directly from the WWW server.

References

Arrigo,P., Giuliano,F., Scalia,F., Rapallo,A. and Damiani,G. (1991) *CABIOS*, **7**, 353–357.
Bairoch,A. and Boeckmann,B. (1994) *Nucleic Acids Res.*, **22**, 3578–3580.
Brunak,S., Engelbrecht,J. and Knudsen,S. (1991) *J. Mol. Biol.*, **220**, 49–65.
Fleischmann,R. *et al.* (1995) *Science*, **269**, 449–604.
Gierasch,L.M. (1989) *Biochemistry*, **28**, 923–930.
Hirst,J.D. and Sternberg,M.J.E. (1992) *Biochemistry*, **31**, 7211–7218.
Hobohm,U., Scharf,M., Schneider,R. and Sander,C. (1992) *Protein Sci.*, **1**, 409–417.
Ladunga,I., Czakó,F., Csabai,I. and Geszti,T. (1991) *CABIOS*, **7**, 485–487.
Mathews,B. (1975) *Biochim. Biophys. Acta*, **405**, 442–451.
Nielsen,H., Engelbrecht,J., von Heijne,G. and Brunak,S. (1996a) *Proteins*, **24**, 165–177.
Nielsen,H., Engelbrecht,J., von Heijne,G. and Brunak,S. (1997) *Int. J. Neural Sys.*, in press.
Presnell,S.R. and Cohen,F.E. (1993) *Annu. Rev. Biophys. Biomol. Struct.*, **22**, 283–298.
Qian,N. and Sejnowski,T.J. (1988) *J. Mol. Biol.*, **202**, 865–884.
Rapoport,T.A. (1992) *Science*, **258**, 931–936.
Rumelhart,D.E., Hinton,G.E. and Williams,R.J. (1986) In Rumelbart,D., McClelland,J. and the PDP Research Groups (eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*. MIT Press, Cambridge, MA, pp. 318–362.
Schneider,G. and Wrede,P. (1993) *J. Mol. Evol.*, **36**, 586–595.
Schneider,T.D. and Stephens,R.M. (1990) *Nucleic Acids Res.*, **18**, 6097–6100.
Shannon,C.E. (1948) *Bell System Technol. J.*, **27**, 379–423, 623–656.
von Heijne,G. (1983) *Eur. J. Biochem.*, **133**, 17–21.
von Heijne,G. (1985) *J. Mol. Biol.*, **184**, 99–105.
von Heijne,G. (1986) *Nucleic Acids Res.*, **14**, 4683–4690.
von Heijne,G. (1988) *Biochim. Biophys. Acta*, **947**, 307–333.
von Heijne,G. (1990) *J. Membrane Biol.*, **115**, 195–201.
von Heijne,G. and Abrahmsén,L. (1989) *FEBS Lett.*, **244**, 439–446.

Received April 19, 1996; revised September 2, 1996; accepted September 12, 1996

		Due Date:	November 27, 2007
Applicant:	W. James JACKSON	**Art Unit:**	1645
		Confirmation No.:	5261
Application No.:	09/677,752	**Examiner:**	FORD, Vanessa L.
Filed:	October 2, 2000	**Docket:**	2479.0050000/EJH/C-K
For:	**Chlamydia Protein, Gene Sequence and Uses Thereof**	**Atty:**	EJH/C-K

When receipt stamp is placed hereon, the USPTO acknowledges receipt of the following documents:

1. SKGF Cover Letter;
2. Credit Card Payment Form (PTO-2038) in the amount of **$930.00** to cover **$ 810.00** Request for Continued Examination (RCE) Fee; and **$ 120.00** Petition for Extension of Time (One month) Fee.
3. Request for Continued Examination (RCE) Transmittal (PTO/SB/30);
4. Petition for Extension of Time (One month) Under 37 C.F.R. § 1.136(a) (PTO/SB/22);
5. Reply Under 37 C.F.R. § 1.114 with Exhibits 1 and 2;
6. Declaration Under 37 C.F.R. § 1.132 and Exhibits A-H;
7. Seventh Supplemental Information Disclosure Statement Under 37 C.F.R. § 1.97(b);
8. Form PTO/SB/08A (1 page) citing seven (7) documents (US2-US8);
9. Form PTO/SB/08B (1 page) citing five (5) documents (NPL6-NPL10);
10. Copies of four (4) cited documents (NPL6-NPL9); and
11. One (1) return postcard.

741918_1.DOC

Mail Stop: RCE

Please Date Stamp and Return to Our Courier

		Due Date:	November 27, 2007
Applicant:	W. James JACKSON	**Art Unit:**	1645
		Confirmation No.:	5261
Application No.:	09/677,752	**Examiner:**	FORD, Vanessa L.
Filed:	October 2, 2000	**Docket:**	2479.0050000/EJH/C-K
For:	**Chlamydia Protein, Gene Sequence and Uses Thereof**	**Atty:**	EJH/C-K

When receipt stamp is placed hereon, the USPTO acknowledges receipt of the following documents:

1. SKGF Cover Letter;
2. Credit Card Payment Form (PTO-2038) in the amount of **$930.00** to cover **$ 810.00** Request for Continued Examination (RCE) Fee; and **$ 120.00** Petition for Extension of Time (One month) Fee.
3. Request for Continued Examination (RCE) Transmittal (PTO/SB/30);
4. Petition for Extension of Time (One month) Under 37 C.F.R. § 1.136(a) (PTO/SB/22);
5. Reply Under 37 C.F.R. § 1.114 with Exhibits 1 and 2;
6. Declaration Under 37 C.F.R. § 1.132 and Exhibits A-H;
7. Seventh Supplemental Information Disclosure Statement Under 37 C.F.R. § 1.97(b);
8. Form PTO/SB/08A (1 page) citing seven (7) documents (US2-US8);
9. Form PTO/SB/08B (1 page) citing five (5) documents (NPL6-NPL10);
10. Copies of four (4) cited documents (NPL6-NPL9); and
11. One (1) return postcard.

741918_1.DOC

Mail Stop: RCE

Please Date Stamp and Return to Our Courier

Equivalent of Form PTO/SB/08A (10-07)
Approved for use through 10/31/2007. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

SEVENTH SUPPLEMENTAL INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 1 of 1

Complete if Known	
Application Number	09/677,752
Filing Date	October 2, 2000
First Named Inventor	JACKSON, W. James
Art Unit	1645
Examiner Name	FORD, Vanessa L.
Attorney Docket Number	2479.0050000/EJH/C-K

U.S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (If Known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	US2	6,559,294	05/06/2003	Griffais *et al.*	
	US3	6,565,856	05/20/2003	Skeiky *et al.*	
	US4	6,642,023	11/04/2003	Jackson *et al.*	
	US5	6,887,843	05/03/2005	Jackson *et al.*	
	US6	2004/0067524 A1	04/08/2004	Jackson *et al.*	
	US7	2004/0137005 A1	07/15/2004	Jackson *et al.*	
	US8	2005/0048557 A1	03/03/2005	Jackson *et al.*	

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3] Number[4] Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear	T[6]

738484v1

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached.

Equivalent of Form PTO/SB/08B (10-07)
Approved for use through 10/31/2007. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

SEVENTH SUPPLEMENTAL INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 1 of 1

Complete if Known	
Application Number	09/677,752
Filing Date	October 2, 2000
First Named Inventor	JACKSON, W. James
Art Unit	1645
Examiner Name	FORD, Vanessa L.
Attorney Docket Number	2479.0050000/EJH/C-K

NON PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.[1]	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published	T[2]
	NPL6	Longbottom, D., *et al.*, "Molecular Cloning and Characterization of the Genes Coding for the Highly Immunogenic Cluster of 90-Kilodalton Envelope Proteins from the *Chlamydia psittaci* Subtype That Causes Abortion in Sheep," *Infect. Immun. 66*:1317-1324, American Society for Microbiology (1998)	
	NPL7	NCBI Entrez, GenBank Report, Accession No. AAC68472, Stephens, R.S., *et al.* (first available 1998)	
	NPL8	Co-pending and commonly-owned Non-Provisional United States Patent Application No. 11/781,199, inventor Jackson, W. James, filed July 20, 2007 (NOT PUBLISHED)	
	NPL9	Co-pending and commonly-owned Non-Provisional United States Patent Application No. 11/781,203, inventor Jackson, W. James, filed July 20, 2007 (NOT PUBLISHED)	
	NPL10	Co-pending and commonly-owned Non-Provisional United States Patent Application No. 08/942,596, inventors Jackson, W. James, *et al.*, filed October 2, 1997 (NOT PUBLISHED)	

738491v1

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] Applicant is to place a check mark here if English language Translation is attached.